



Summary statistics

- 10A Calculating the mean
- 10B Standard deviation
- 10C Median and mode
- 10D Best summary statistics



Syllabus reference

Data analysis 4

- Summary statistics

In this chapter we are going to further analyse data, completing our understanding of summary statistics; that is, information that summarises a data set with a single number.

ARE YOU READY?

Try the questions below. If you have difficulty with any of them, extra help can be obtained by completing the matching SkillsHEET. Either click on the SkillsHEET icon next to the question on the *Maths Quest Preliminary Course* eBookPLUS or ask your teacher for a copy.

eBookplus

Digital doc

SkillsHEET 10.1
doc-1581

Finding the mean of a list of scores

Finding the mean of a list of scores

1 Find the average of each of the following sets of scores.

- a 1, 3, 4, 6, 8
- b 1.5, 1.2, 1.3, 1.5, 1.8, 1.1, 1.2, 1.7
- c 180, 45, 92, 84, 96, 2, 104, 32, 8, 111

eBookplus

Digital doc

SkillsHEET 10.2
doc-1582

Presenting data as a dot plot

Presenting data as a dot plot

2 Draw a dot plot to represent the following data.

6 8 7 9 4 6 7 8 3 5 7

eBookplus

Digital doc

SkillsHEET 10.3
doc-1583

Presenting data in a frequency table

Presenting data in a frequency table

3 a Display the following sets of scores in a frequency table.

15 16 18 19 15 13 14 13 12 18
15 19 18 12 14 13 17 18 14 16

b Use the classes 0–9, 10–19, 20–29, ... etc. to display the scores below in a frequency table.

45 13 9 12 28 19 36 37 28 42
28 18 39 28 36 40 28 37 28 48

eBookplus

Digital doc

SkillsHEET 10.5
doc-1585

Presenting data as a stem-and-leaf plot

Presenting data as a stem-and-leaf plot

4 Place the scores shown below in a stem-and-leaf plot.

48 31 20 20 46 20 25 41 32 49
24 31 31 28 46 48 41 46 27 46
29 24 36 44 29 40 41 20 39 41

10A Calculating the mean

INVESTIGATE: Average — what does it mean?

Survey a group of people about what they believe is meant by the word ‘average’. Use their answers to describe what the word is generally understood to mean.



When looking at a set of statistics we are often asked for the average. The average is a figure that describes a typical score. In statistics, the correct term for the average is the **mean**. The mean is the first of three measures of **central tendency** that we will be studying. The others are the median and the mode.

The statistical symbol for the mean is \bar{x} . The formula for the mean is $\bar{x} = \frac{\sum x}{n}$.

In Mathematics, the symbol Σ (sigma) means sum or total, x represents each individual score in a list and Σx is therefore the sum of the scores. The sum is divided by n , which represents the number of scores.

WORKED EXAMPLE 1

Find the mean of the scores 17, 16, 13, 15, 16, 20, 10, 15.

THINK

Method 1

- 1 Find the total of all scores.
- 2 Divide the total by 8 (the number of scores).

Method 2

- 1 From the **MENU** select **STAT**.

WRITE

$$\begin{aligned}\text{Total} &= 17 + 16 + 13 + 15 + 16 + 20 + 10 + 15 \\ &= 122\end{aligned}$$

$$\begin{aligned}\text{Mean} &= \frac{122}{8} \\ &= 15.25\end{aligned}$$



- 2 Delete any existing data and enter the scores into **List 1**.

List 1	List 2	List 3	List 4
14			
16			
13			
15			
16			

- 3 Press **F2** for **CALC**, then **F6** for **SET**. For **1Var Xlist**, enter **List 1** by pressing **F1**. This means that the scores are stored in List 1. For **1Var Freq**, enter **1** by pressing **F1**. This means every score entered has a frequency of 1.

1Var Xlist	:List1
1Var Freq	:1
2Var Xlist	:List1
2Var Ylist	:List1
2Var Freq	:List1

- 4 Press **EXIT** to return to the previous screen. Press **F1** for **1Var** to display all summary statistics. The mean (\bar{x}) is the first summary statistic displayed.

1-Variable	
\bar{x}	=15.25
s_x	=2.50
s_x^2	=6.25
$\bar{x}n$	=2.72717802
$\bar{x}n-1$	=2.91547594
n	=8

As we have seen, large amounts of data are often presented in a frequency table. To calculate the mean in such a case, we need to add an extra column to the table. This column is the $f \times x$ column. In this column, we multiply each frequency by the score. We then total this column to find the total of all scores and divide this by the sum of the frequency column. Written as a formula this is:

$$\bar{x} = \frac{\sum f \times x}{\sum f}$$

WORKED EXAMPLE 2

Complete the frequency table at right, then calculate the mean.

Score (x)	Frequency (f)	$f \times x$
4	3	
5	7	
6	11	
7	13	
8	10	
9	6	
$\Sigma f =$		$\Sigma f \times x =$

THINK

Method 1

- Complete the $f \times x$ column by multiplying each score by the frequency.
- Sum the frequency and $f \times x$ columns.

WRITE

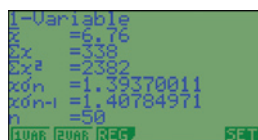
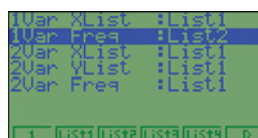
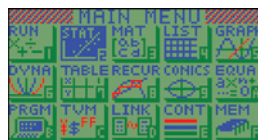
Score (x)	Frequency (f)	$f \times x$
4	3	12
5	7	35
6	11	66
7	13	91
8	10	80
9	6	54
$\Sigma f = 50$		$\Sigma f \times x = 338$

- 3 Use the formula to calculate the mean.

$$\begin{aligned}\bar{x} &= \frac{\sum f \times x}{\sum f} \\ &= \frac{338}{50} \\ &= 6.76\end{aligned}$$

Method 2

- 1 From the **MENU** select **STAT**.
- 2 Delete any existing data and enter the scores in List 1 and the frequencies in List 2.
- 3 Press **F2** for **CALC**, then **F6** for **SET**. For **1Var Xlist**, enter **List 1** by pressing **F1**. For **1Var Freq**, enter **List 2** by pressing **F3**. This means the entries in List 2 are the frequencies corresponding to the entries in List 1.
- 4 Press **EXIT** to return to the previous screen. Press **F1** for **1Var** to display all summary statistics. The mean (\bar{x}) is the first summary statistic displayed.



The same method is used when the frequency table is given in terms of grouped data. In these cases, however, to calculate the $f \times x$ column we use the class centre multiplied by the frequency. In these cases, we obtain an estimate of the mean rather than an exact mean.

WORKED EXAMPLE 3

Complete the frequency distribution table and use it to estimate the mean of the distribution.

Class	Class centre (x)	Frequency (f)	$f \times x$
25–29		4	
30–34		9	
35–39		13	
40–44		12	
45–50		7	
		$\Sigma f =$	$\Sigma f \times x =$

THINK

- 1 Calculate the class centres.
- 2 Multiply each class centre by the frequency to complete the $f \times x$ column.

WRITE

eBookplus

Tutorial
int-2327

Worked example 3

- 3 Sum the frequency and the $f \times x$ column.

Class	Class centre (x)	Frequency (f)	$f \times x$
25–29	27	4	108
30–34	32	9	288
35–39	37	13	481
40–44	42	12	504
45–50	47	7	329
		$\Sigma f = 45$	$\Sigma f \times x = 1710$

- 4 Use the formula to calculate the mean.

$$\begin{aligned}\bar{x} &= \frac{\Sigma f \times x}{\Sigma f} \\ &= \frac{1710}{45} \\ &= 38\end{aligned}$$

In most cases, when calculating the mean you will use your calculator and will need to set it to statistics mode. Once this is done, each score is entered and the **M+** function pressed. When all scores are entered, the mean is found by using the \bar{x} function.

If the data are presented in the form of a frequency distribution table, you will need to check how to enter multiple scores. On many calculators, you press score \times frequency followed by **M+**, but check with your teacher as to how your calculator works.

For all statistical questions, when using your calculator clear the memory at the beginning of each question. Most calculators will display the number of scores you have entered after each entry. This is a useful check that you have cleared the memory and entered the data correctly.

WORKED EXAMPLE 4

Use your calculator to find the mean of:

a 10, 15, 47, 23, 56

b

Score	Frequency
67	10
68	23
69	35
70	28
71	12



THINK

- a
- Put your calculator on to statistics mode and clear the memory.
 - Press each score followed by **M+**.
 - Get the mean by pressing \bar{x} .
- b
- Set your calculator to statistics mode and clear the memory.
 - Press each score \times frequency then **M+**.
 - Get the mean by pressing \bar{x} .

WRITE

a

$$\text{Mean} = 30.2$$

b

$$\text{Mean} = 69.1$$

REMEMBER

1. The mean is the statistical term for average.
2. The mean is calculated by adding all scores then dividing by the number of scores.
3. When calculating the mean from a frequency distribution table, a column for frequency \times score ($f \times x$) is added. The mean is then calculated using the formula:
$$\bar{x} = \frac{\sum f \times x}{\sum f}$$
4. If the frequency distribution uses grouped data, the $f \times x$ column is calculated using class centres.
5. The mean can be calculated using your calculator. To do so, set the calculator to statistics mode, enter the scores using the **M+** function and make sure you know how to retrieve the mean using the \bar{x} function.

EXERCISE

10A

Calculating the mean

eBookplus

Digital doc

SkillSHEET 10.1
doc-1581

Finding the
mean of a list
of scores

- 1 **WE1** Calculate the mean of each of the following sets of scores.

- a 4, 8, 3, 5, 5
- b 16, 24, 30, 35, 23, 11, 45, 28
- c 65, 92, 56, 84
- d 9.2, 9.7, 8.8, 8.1, 5.6, 7.5, 8.5, 6.4, 7.0, 6.4
- e 356, 457, 182, 316, 432, 611, 299, 355

- 2 Majid sits for five tests in Mathematics. His marks on the tests were 45%, 90%, 67%, 86% and 75%. Calculate Majid's mean mark on the five tests.

- 3 An oil company surveys the price of petrol in eight Sydney suburbs. The results are below.

Manly	132.9 c/L
Cronulla	129.9 c/L
Wentworthville	125.5 c/L
Campbelltown	125.9 c/L
Lakemba	121.9 c/L
Liverpool	119.9 c/L
Epping	128.9 c/L
Penrith	120.9 c/L

Based on these results, calculate the mean price of petrol in cents per litre in Sydney.

- 4 The seven players on a netball team have the following heights: 1.65 m, 1.81 m, 1.75 m, 1.78 m, 1.88 m, 1.92 m and 1.86 m. Calculate the mean height of the players on this team, correct to 2 decimal places.

- 5 A golf ball manufacturer randomly tests the mass of 10 golf balls from a batch. The batch will be considered satisfactory if the average mass of the balls is between 44.8 g and 45.2 g. The mass, in grams, of those tested are:

45.19, 45.06, 45.35, 44.78, 45.47, 44.68, 44.95, 45.32, 44.60, 44.95.

Will the batch be passed as satisfactory?

eBookplus

Digital doc

SkillSHEET 10.2
doc-1582

Presenting

data as a
dot plot

eBookplus

Digital doc

SkillSHEET 10.3
doc-1583

Presenting

data in a
frequency
table



eBookplus**Digital doc**SkillSHEET 10.4
doc-1584**Organising
data into class
intervals****eBookplus****Digital doc**SkillSHEET 10.5
doc-1585**Presenting
data as a
stem-and-leaf
plot****eBookplus****Digital doc**EXCEL
Spreadsheet
doc-1586
Mean**eBookplus****Digital doc**EXCEL
Spreadsheet
doc-1587
Mean (DIY)

- 6 WE2**
- The marks out of 10 on a spelling test are recorded in the frequency table below.

a Copy and complete the table.**b** Use the formula: $\bar{x} = \frac{\sum f \times x}{\sum f}$ to calculate the mean.

Score	Frequency	$f \times x$
4	2	
5	4	
6	5	
7	9	
8	3	
9	5	
10	2	
$\Sigma f =$		$\Sigma f \times x =$

- 7**
- An electrical store records the number of televisions sold each week over a year. The results are shown in the table below.

No. of televisions sold	No. of weeks	$f \times x$
16	4	
17	4	
18	3	
19	6	
20	7	
21	12	
22	8	
23	2	
24	4	
25	2	
$\Sigma f =$		$\Sigma f \times x =$



- a** Copy and complete the table.
- b** Calculate the mean number of televisions sold each week over the year. Give your answer correct to 1 decimal place.
- 8** In a soccer season a team played 50 matches. The number of goals scored in each match is shown in the table below.

No. of goals	0	1	2	3	4	5
No. of matches	4	9	18	10	5	4

- a** Redraw this table in the form of a frequency distribution table.
- b** Use your table to calculate the mean number of goals scored each game.
- 9** A clothing store records the dress sizes sold during a day. The results are shown below.

12 14 10 12 8 12 16 10 8 12
 10 12 18 10 12 14 16 10 12 12
 12 14 18 10 14 12 12 14 14 10

- a** Present this information in a frequency table.
- b** Calculate the mean dress size sold this day.

- 10 MC** There are eight players in a Rugby forward pack. The mean mass of the players is 104 kg. The total mass of the forward pack is:

A 13 kg **B** 104 kg **C** 112 kg **D** 832 kg

- 11 MC** A small business employs five people on a mean wage of \$580 per week. A manager is then employed and receives \$700 per week. What is the mean wage of the six employees?

A \$580
B \$600
C \$680
D \$3600

- 12 MC** The mean height of five starting players in a basketball match is 1.82 m. During a time out, a player who is 1.78 m tall is replaced by a player 1.88 m tall. What is the mean height of the players after the replacement has been made?

A 1.78 m
B 1.82 m
C 1.84 m
D 1.88 m



- 13 WE3** The table below shows a set of class marks on a test out of 100.

Class	Class centre (x)	Frequency (f)	$f \times x$
31–40		1	
41–50		3	
51–60		4	
61–70		7	
71–80		11	
81–90		2	
91–100		2	
		$\Sigma f =$	$\Sigma f \times x =$

- a** Copy and complete the frequency distribution table.
b Use the table to calculate the mean class mark.

- 14** In the heats of the 100 m freestyle at a swimming meet, the times of the swimmers were recorded in the table below.

Time	Class centre	No. of swimmers	$f \times x$
50.01–51.00		4	
51.01–52.00		12	
52.01–53.00		23	
53.01–54.00		38	
54.01–55.00		15	
55.01–56.00		3	
		$\Sigma f =$	$\Sigma f \times x =$

- a** Copy and complete the frequency distribution table.
b Use the table to calculate the mean time.

- 15** A cricketer played 50 innings in test cricket for the following scores.

23 65 8 112 54 0 84 12 21 4
 25 105 74 40 1 15 33 45 21 47
 16 70 22 33 21 8 34 36 5 7
 69 104 57 78 158 0 51 16 6 16
 0 49 0 14 28 52 21 3 3 7

- a** Put the above information into a frequency distribution table using appropriate groupings.
b Use the table to estimate the batting average for this player.



- 16** **WE4** Use the statistics function on your calculator to find the mean of each of the following scores, correct to 1 decimal place.

- a** 11, 15, 13, 12, 21, 19, 8, 14
b 2.8, 2.3, 3.6, 2.9, 4.5, 4.2
c 41, 41, 41, 42, 43, 45, 45, 45, 45, 46, 49, 50

- 17** Use your calculator to find the mean from each of the following tables.

a

Score	Frequency
3	7
4	10
5	18
6	19
7	38
8	27
9	10
10	5

b

Score	Frequency
28	5
29	18
30	25
31	25
32	14
33	10
34	3

- 18** The table below shows the heights of a group of people.

Height (cm)	Class centre	Frequency
150–154	152	7
155–159	157	14
160–164	162	13
165–169	167	23
170–174	172	24
175–179	177	12

Calculate the mean of this distribution.

- 19** Seventy students were timed on a 100 m sprint during their P.E. class. The results are shown in the table below.

Time (s)	12–13	13–14	14–15	15–16	16–17
Number	13	17	25	15	10

- Calculate the class centre for each group in the distribution.
 - Use your calculator to find the mean of the distribution.
- 20** A drink machine is installed near a quiet beach. The number of cans sold each day over the first 10 weeks after its installation is shown below.

4 39 31 31 50 43 70 45 57 71 18 26 3 52
 51 59 33 51 27 62 30 90 3 30 97 59 33 44
 99 62 72 6 42 83 19 49 11 6 63 4 53 20
 45 58 1 9 79 41 2 33 97 71 52 97 69 83
 39 84 92 43 71 98 8 97 18 89 21 9 4 17

- Put this information into a frequency distribution table using the classes 1–10, 11–20, 21–30 etc.
- Calculate the mean number of cans sold per day over these 10 weeks.



Further development

- 21** The mean of 5 scores is 12.6.
- What is the total of the five scores?
 - An extra score of 19.2 is added to the data set. What is the new total of the scores?
 - Find the mean of the six scores.
- 22** The mean of 9 scores is 58. A tenth score of 19 is added to the data set. Find the new mean of the data set.
- 23** The data below shows the ages of 10 people working out at the gym.
- 23 24 19 59 23 22 16 18 25
- Find the mean of the data set.
 - One score which is vastly different to all other scores is called an outlier. What is the outlier in this data set?
 - Calculate the mean of the nine remaining scores with the outlier omitted.
 - Write a sentence describing the effect that the outlier has on the mean.
- 24** The mean of a data set containing 8 scores is 63. After a ninth score is added the mean falls to 60. What was the ninth score?
- 25** Livinia has an average score of 14 for 6 essays that she has written. What score will she need to achieve for her next essay in order to lift her mean to 15.5?
- 26** Describe the effect on the mean if a score:
- greater than the mean is added to the data set
 - less than the mean is added to the data set.

10B Standard deviation

eBookplus

Interactivity

int-2402

Standard

deviation

In the previous chapter, we discussed using the range and the interquartile range as a measure of the spread of a data set. The most commonly used measure of spread is the **standard deviation**.

The standard deviation is a measure of how much a typical score in a data set differs from the mean.

The standard deviation is found by entering a set of scores into your calculator, just as you do when you are finding the mean. Your calculator will have a function that gives the standard deviation.

There are two standard deviation functions on your calculator. The first, σ_n , is the population standard deviation. This function is used when the statistical analysis is conducted on the entire population.

WORKED EXAMPLE 5

Below are the scores out of 100 by a class of 20 students on a Science exam. Calculate the mean and the standard deviation.

87 69 95 73 88 47 95 63 91 66
59 70 67 83 71 57 82 65 84 69

THINK

Method 1

- 1 Enter the data set into your calculator.
- 2 Retrieve the mean using the \bar{x} function.
- 3 Retrieve the standard deviation using the σ_n function.

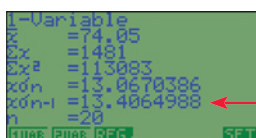
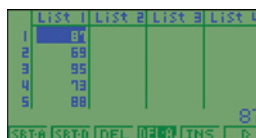
WRITE

$$\bar{x} = 74.05$$

$$\sigma_n = 13.07$$

Method 2

- 1 From the **MENU** select **STAT**.
- 2 Delete any existing data and enter the scores into **List 1**.
- 3 Press **F2** for **CALC**, then **F6** for **SET**. Set the calculator up for a list of scores as shown earlier and as shown at right.
- 4 Press **EXIT** to return to the previous screen, then **F1** to display the summary statistics. The population standard deviation is displayed by the symbol σ_n .



Population standard deviation

When the statistical analysis is done using a sample of the population, a slightly different standard deviation function is used. Called the sample standard deviation, this value will be slightly higher than the population standard deviation. The sample standard deviation will be found on your calculator using the σ_{n-1} or the s_n function.

WORKED EXAMPLE 6

Ian surveys twenty Year 11 students and asks how much money they earn from part-time work each week. The results are given below.

\$65	\$82	\$47	\$78	\$108	\$94	\$60	\$79	\$88	\$91
\$50	\$73	\$68	\$95	\$83	\$76	\$79	\$72	\$69	\$97

Calculate the mean and standard deviation.

THINK

Method 1

- 1 Enter the statistics into your calculator.
- 2 Retrieve the mean using the \bar{x} function.
- 3 Retrieve the standard deviation using the σ_{n-1} function, as a sample has been used.

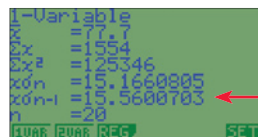
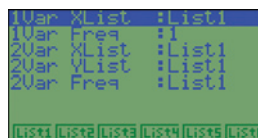
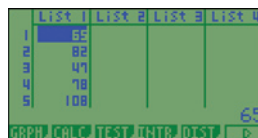
WRITE

$$\bar{x} = \$77.70$$

$$\sigma_{n-1} = \$15.56$$

Method 2

- 1 From the **MENU** select **STAT**.
- 2 Delete any existing data and enter the scores into **List 1**.
- 3 Press **F2** for **CALC**, then **F6** for **SET**. Set the calculator up for a list of scores as shown earlier and as shown at right.
- 4 Press **EXIT** to return to the previous screen, then **F1** to display the summary statistics. The sample standard deviation is displayed by the symbol $x\sigma_{n-1}$.



For most examples, you will need to read the question carefully to decide whether to use the population or the sample standard deviation.

The standard deviation can also be calculated when the data are presented in table form. This is done by entering the data in the same way as they were when calculating the mean earlier in this chapter.

WORKED EXAMPLE 7

The table below shows the scores of a class of thirty Year 3 students on a spelling test.

Score	Frequency
4	1
5	2
6	4
7	9
8	6
9	7
10	1



Calculate the mean and standard deviation.

THINK

- 1 Enter the data into your calculator using score \times frequency.
- 2 Retrieve the mean using the \bar{x} function.
- 3 Retrieve the standard deviation using the σ_n function, as the whole population is included in the statistics.

WRITE

$$\bar{x} = 7.4$$

$$\sigma_n = 1.4$$

Once we have calculated the standard deviation, we can make conclusions about the reliability and consistency of the data set. The lower the standard deviation, the less spread out the data set is. By using the standard deviation, we can determine whether a set of scores is more or less consistent (or reliable) than another set. The standard deviation is the best measure of this because, unlike the range or interquartile range as a measure of dispersion, the standard deviation considers the distance of every score from the mean.

A higher standard deviation means that scores are less clustered around the mean and less dependable. For example, consider the following two students:

$$\begin{array}{ll} \text{Student A: } \bar{x} = 60 & \sigma_n = 5 \\ \text{Student B: } \bar{x} = 60 & \sigma_n = 15 \end{array}$$

Both students have the same mean. However, student A has a standard deviation of 5 and student B has a standard deviation of 15. Student A is far more consistent and can confidently be expected to score around 60 in any future exam. Student B is more inconsistent but is probably capable of scoring a higher mark than student A. This concept will be discussed further during the HSC course.

WORKED EXAMPLE 8

Two brands of light globe are tested to see how long they will burn (in hours).

Brand X:	850	950	1400	875	1200
	1150	1000	900	850	825
Brand Y:	975	1100	1050	1000	975
	950	1075	1025	950	900

Which of the two brands of light globe is more reliable?



THINK

- 1 Enter both sets of data into your calculator.
- 2 Choose the sample standard deviation because a sample of each light globe brand has been chosen.
- 3 Write down the sample standard deviation for each brand.
- 4 The brand with the lower standard deviation is the more reliable.

WRITE

Brand X: $\sigma_{n-1} = 190.4$

Brand Y: $\sigma_{n-1} = 62.4$

Brand Y is the more reliable as it has a lower standard deviation.

REMEMBER

1. The standard deviation is a measure of the spread of a data set.
2. Standard deviation is found on your calculator by entering the data set using the calculator's statistical mode.
3. The population standard deviation is used when an entire population is considered in the statistical analysis and can be found on the calculator using the σ_n function.
4. The sample standard deviation is used when a sample of the population is used in the analysis and can be found using the σ_{n-1} function.

EXERCISE**10B****Standard deviation**

- 1 **WE5** For each of the sets of scores below, calculate the standard deviation. Assume that the scores represent an entire population and answer correct to 2 decimal places.

a 3, 5, 8, 2, 7, 1, 6, 5	b 11, 8, 7, 12, 10, 11, 14
c 25, 15, 78, 35, 56, 41, 17, 24	d 5.2, 4.7, 5.1, 12.6, 4.8
e 114, 12, 3.6, 42.8, 0.5	
- 2 **WE6** For each of the sets of scores below, calculate the sample standard deviation, correct to 2 decimal places.

a 25, 36, 75, 85, 6, 49, 77, 80, 37, 66
b 4.8, 9.3, 7.1, 9.9, 7.0, 4.1, 6.2
c 112, 25, 56, 81, 0, 5, 178, 99, 41
d 0.3, 0.3, 0.3, 0.4, 0.5, 0.6, 0.8, 0.8, 0.8, 0.9, 1.0
e 56, 1, 258, 45, 23, 58, 48, 35, 246
- 3 For each of the following, state whether it is appropriate to use the population standard deviation or the sample standard deviation.

a A quality control officer tests the life of 50 batteries from a batch of 1000.
b The weight of every bag of potatoes is checked and recorded before being sold.
c The number of people who attend every football match over a season is analysed.
d A survey of 100 homes records the number of cars in each household.
e The score of every HSC student in Mathematics is recorded.



- 4 The band 'Aquatron' is to release a new CD. The recording company needs to predict the number of copies that will be sold at various music stores throughout Australia. To do so, a sample of 10 music stores supplied information about the sales of the previous CD released by Aquatron, as shown below.

580 695 547 236 458 620 872 364 587 1207

- Calculate the mean number of sales at each store.
 - Should the population or sample standard deviation be used in this case?
 - What is the value of the appropriate standard deviation?
- 5 A supermarket chain is analysing its sales over a week. The chain has 15 stores and the sales for each store for the past week were (in \$million):

1.5 2.1 2.4 1.8 1.1 0.8 0.9 1.1 1.4 1.6 2.0 0.7 1.2 1.7 1.3

- Calculate the mean sales for the week.
 - Should the population or sample standard deviation be used in this case?
 - What is the value of the appropriate standard deviation?
- 6 **WE7** Use the statistical function on your calculator to find the mean and standard deviation (correct to 1 decimal place) for the information presented in the following tables. In each case, use the population standard deviation.

a

Score	Frequency
3	12
4	24
5	47
6	21
7	7

b

Score	Frequency
45	1
46	16
47	39
48	61
49	52
50	36

c

Score	Frequency
75	22
76	17
77	8
78	10
79	12
80	21
81	29

- 7 Copy and complete the class centre column for each of the following distributions and use your calculator to find an estimate for the mean and standard deviation (correct to 2 decimal places). In each case use the population standard deviation.

a

Class	Class centre	Frequency
10–12		12
13–15		16
16–18		25
19–21		28
22–24		13

b

Class	Class centre	Frequency
31–40		15
41–50		28
51–60		36
61–70		19
71–80		8
81–90		7
91–100		2

c

Class	Class centre	Frequency
0–4		15
5–9		24
10–14		31
15–19		33
20–24		29
25–29		17

- 8 **WE8** Below are the marks achieved by two students in five tests.

Brianna: 75, 80, 70, 72, 78

Katie: 50, 95, 90, 80, 55

- a Calculate the mean and standard deviation for each student.
b Which of the two students is more consistent? Explain your answer.
- 9 **MC** From Year 11, 21 students are chosen to complete a test. The scores are shown in the table below.

Class	Frequency
10–20	1
20–30	6
30–40	9
40–50	4
50–60	1

When preparing an analysis of the typical performance of Year 11 students on the test, the standard deviation used is:

- A 9.209
B 9.437
C 21
D 34.048
- 10 **MC** The results below are Ian's marks in four exams for each subject that he studies.

English: 63 85 78 50

Maths: 69 71 32 97

Biology: 45 52 60 41

Geography: 65 78 59 61

In which subject does Ian achieve the most consistent results?

- A English
B Maths
C Biology
D Geography
- 11 The following frequency distribution gives the prices paid by a car wrecking yard for a sample of 40 car wrecks.

Price (\$)	Frequency
0–500	2
500–1000	4
1000–1500	8
1500–2000	10
2000–2500	7
2500–3000	6
3000–3500	3

Find the mean and standard deviation of the price paid for these wrecks.



- 12** The table below shows the life of a sample of 175 household light globes.

Life (hours)	Frequency
200–250	2
250–300	5
300–350	12
350–400	25
400–450	42
450–500	38
500–550	26
550–600	15
600–650	7
650–700	3



- Find the range of the data.
- Use the class centres to find the mean and standard deviation in the lifetimes of this sample of light globes.

- 13** Crunch and Crinkle are two brands of potato crisps. Each are sold in packets nominally of the same size and for the same price. Upon investigation of a sample of packets of each, it is found that Crunch and Crinkle have the same mean mass (25 g). The standard deviation of the masses of Crunch is, however, 5 g and the standard deviation of the masses of Crinkle is 2 g. Which brand do you think represents better value for money under these circumstances? Why?

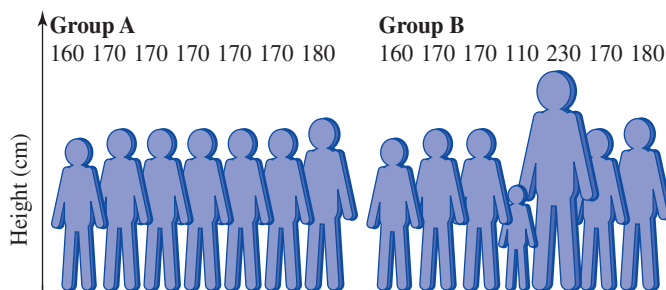


eBookplus

Digital doc
WorkSHEET 10.1
doc-2484

Further development

- 14** Consider the following two groups of people.



- Calculate the mean height of each group.
- Are the groups really the same?

- c In which group would you expect the greater standard deviation?
- d Calculate the standard deviation to confirm your answer.

15 Consider the set of scores 3, 5, 8, 2, 7, 4, 5, 6.

- a Find the mean of the data set.
- b Find the standard deviation of the data set.
- c A score of 9 is added to the data set. What is the difference between this score and the mean?
- d What is the standard deviation of the data set once the extra score is added?

16 Consider the data set 25, 15, 78, 35, 56, 41, 17, 21.

- a Find the mean and standard deviation of the data set.
- b An extra score is added to the data set. Copy and complete the table below to explore the effect that adding an extra score has on the standard deviation.

Extra score	Difference from mean	Standard deviation after score added	Standard deviation (increase or decrease)
8			
30			
90			
50			

17 A data set has a mean of 48 and a standard deviation of 23. A score of 55 is added to the data set.

- a What effect will adding the extra score have on the mean? Explain your answer.
- b What effect will adding the extra score have on the standard deviation? Explain your answer.

18 **MC** A data set has a mean of 36 and a standard deviation of 8. A score of 12 is added to the data set. What will be the effect on the mean and the standard deviation?

- A The mean will decrease and the standard deviation will decrease.
- B The mean will decrease and the standard deviation will increase.
- C The mean will increase and the standard deviation will decrease.
- D The mean will increase and the standard deviation will increase.

19 Describe in your own words, how adding an extra score to a data set will affect the standard deviation.

10C Median and mode

So far we have used the mean as a measure of the typical score in a data set. Consider the case of someone who is analysing the typical house price in an area. On a particular day, five houses are sold in the area for the following prices:

\$375 000 \$349 000 \$360 000 \$411 000 \$1 250 000

For these five houses the mean price is \$549 000. The mean is much greater than most of the houses in the data set. This is because there is one score that is much greater than all the others. For such data sets, we need to use a different measure of central tendency.

In the previous chapter, we introduced the median as the middle score in a data set, when all scores are arranged in order. For the above data set, the median house price is \$375 000, a much better measure of the typical house price in this area.

WORKED EXAMPLE 9

Calculate the median of the scores 3, 5, 8, 4, 4, 6, 9, 1, 6.

THINK

- 1 Rewrite the scores in ascending order.
- 2 The median is the middle score.

WRITE

1, 3, 4, 4, 5, 6, 6, 8, 9
 Median = 5

The median becomes more complicated when there is an even number of scores because there are two scores in the middle. When there is an even number of scores, the median is the average of the two middle scores.

WORKED EXAMPLE 10

Find the median of the scores 13, 13, 16, 12, 19, 18, 20, 18.

THINK

- 1 Write the scores in ascending order.
- 2 There is an even number (8) scores, so average the two middle scores.

WRITE

12, 13, 13, 16, 18, 18, 19, 20
 Median = $\frac{16 + 18}{2}$
 = 17

The median can also be calculated from the cumulative frequency column of a frequency table.

The cumulative frequency column puts the scores into order and tells us what score is in each position. Consider the frequency distribution table below.

Score	Frequency	Cumulative frequency	
4	1	1	The 1st score is 4.
5	6	7	The 2nd–7th scores are 5.
6	9	16	The 8th–16th scores are 6.
7	8	24	The 17th–24th scores are 7.
8	4	28	The 25th–28th scores are 8.
9	2	30	The 29th and 30th scores are 9.

There are 30 scores in this distribution and so the middle two scores will be the 15th and 16th scores. By looking down the cumulative frequency column we can see that these scores are both 6. Therefore, 6 is the median of this distribution.

WORKED EXAMPLE 11

Find the median for the frequency distribution at right.

Score	Frequency
34	3
35	8
36	12
37	9
38	8
39	5

THINK

Method 1

- 1 Redraw the frequency table with a cumulative frequency column.
- 2 There are 45 scores and so the middle score is the 23rd score.
- 3 Look down the cumulative frequency column to see that the 23rd score is 36.

Method 2

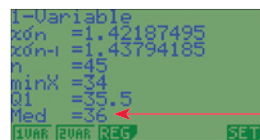
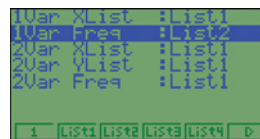
- 1 From the **MENU** select **STAT**.
- 2 Delete any existing data and enter the scores in **List 1** and the frequencies in **List 2**.
- 3 Press **F2** for **CALC**, then **F6** for **SET**. Set the calculator up for data stored in a frequency table as shown earlier in the chapter and as shown by the screen at right.
- 4 Press **EXIT** to return to the previous screen, then **F1** to display the summary statistics. To see the median you will need to use the arrow keys to scroll down the screen by three lines.

WRITE

Score	Frequency	Cumulative frequency
34	3	3
35	8	11
36	12	23
37	9	32
38	8	40
39	5	45

Median = 23rd score

$$= 36$$



Median

When the frequency table presents grouped data, the median is estimated from the ogive as shown in the previous chapter.

There are many examples where neither the mean nor the median is the appropriate measure of the typical score in a data set.

Consider the case of a clothing store. It needs to re-order a supply of dresses. To know what sizes to order it looks at past sales of this particular style and gathers the following data:

8 12 14 12 16 10 12 14 16 18
14 12 14 12 12 8 18 16 12 14

For this data set the mean dress size is 13.2. Dresses are not sold in size 13.2, so this has very little meaning. The median is 13, which also has little meaning as dresses are sold only in even-numbered sizes.

What is most important to the clothing store is the dress size that sells the most. In this case size 12 occurs most frequently. The score that has the highest frequency is called the **mode**.

WORKED EXAMPLE 12

Find the mode of the scores below.

4, 5, 9, 4, 6, 8, 4, 8, 7, 6, 5, 4

THINK

Method 1

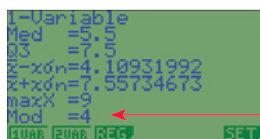
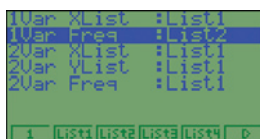
The score 4 occurs most often and so it is the mode.

Method 2

- From the **MENU** select **STAT**.
- Delete any existing data and enter the scores into **List 1**.
- Press **F2** for **CALC**, then **F6** for **SET**. Set the calculator up for a list of scores as shown earlier and as shown at right.
- Press **EXIT** to return to the previous screen, then **F1** to display the summary statistics. To see the mode you will need to use the arrow keys to scroll down to the last line of the display.

WRITE

Mode = 4



Mode

When two scores occur most often an equal number of times, both scores are given as the mode. In this situation the scores are **bimodal**. If all scores occur an equal number of times, then the distribution has no mode. The Casio CFX-9850 shows only the highest mode. The Casio FX-9860GAU shows all modes, as well as the number of modes and the frequency of each.

To find the mode from a frequency distribution table, we simply give the score that has the highest frequency.

WORKED EXAMPLE 13

For the frequency distribution below, state the mode.

Score	14	15	16	17	18	19
Frequency	3	6	11	14	10	7

THINK

The highest frequency is 14, which belongs to the score 17 and so 17 is the mode.

WRITE

Mode = 17

When a table is presented using grouped data, we do not have a single mode. In these cases, the class with the highest frequency is called the modal class.

REMEMBER

1. The median is the middle score in a data set or the average of the two middle scores.
2. The median can be found using the cumulative frequency column of a frequency table.
3. The mode is the score that occurs the most.

EXERCISE

10C

Median and mode

eBookplus

Digital doc

EXCEL Spreadsheet
doc-1588

Median

- 1 **WE9** The scores of seven people on a spelling test are given below.

5 6 5 8 5 9 8

Calculate the median of these marks.

- 2 **WE10** Below are the scores of eight people who played a round of golf.

75 80 81 76 84 83 81 82

Calculate the median for this set of scores.

- 3 Find the median for each of the following sets of scores.

a 3, 4, 5, 5, 5, 6, 9

b 5.6, 5.2, 5.4, 5.3, 5.8, 5.4, 5.3, 5.4

c 45, 62, 39, 88, 75

d 102, 99, 106, 108, 101, 103, 102, 105, 102, 101

- 4 A factory has 80 employees. Over a two-week period the number of people absent from work each day was recorded and the results are shown below.

3, 1, 5, 4, 3, 25, 4, 2, 4, 5

a Calculate the median number of people absent from work each day.

b Calculate the mean number of people absent from work each day.

c Does the mean or the median give a better measure of the typical number of people absent from work each day? Explain your answer.

eBookplus

Digital doc

EXCEL Spreadsheet
doc-1589

Median (DIY)

eBookplus

Digital doc

EXCEL Spreadsheet
doc-1590

Mode

eBookplus

Digital doc

EXCEL Spreadsheet
doc-1591

Mode (DIY)



- 5 WE11** The table at right shows the number of cans of drink sold from a vending machine at a high school each day.

- a** Copy and complete the frequency distribution table.
b Use the table to calculate the median number of cans of drink sold each day from the vending machine.

Score	Frequency	Cumulative frequency
17	4	
18	9	
19	6	
20	12	
21	8	
22	5	
23	4	
24	2	

- 6** The table at right shows the number of accidents a tow truck attends each day over a three-week period. Calculate the median number of accidents attended to by the tow truck each day.

No. of accidents	No. of days
2	4
3	12
4	3
5	1
6	1

- 7** The table at right shows the number of errors made by a machine each day over a 50-day period. Calculate the median number of errors made by the machine each day.



No. of errors per day	Frequency
0	9
1	18
2	13
3	6
4	3
5	1

- 8 MC** There are 25 scores in a distribution. The median score will be the:

- A** 12th score
B 12.5th score
C 13th score
D average of the 12th and 13th score.

- 9 MC** For the scores 4, 5, 5, 6, 7, 7, 9, 10 the median is:

- A** 5 **B** 6 **C** 6.5 **D** 7

- 10 MC** Consider the frequency table at right. The median of these scores is:

- A** 2
B 3
C 8
D 13

Score	Frequency
1	12
2	13
3	8
4	7
5	5

- 11** The frequency distribution table below shows the number of sick days taken by each worker in a small business.

Days sickness	Frequency	Cumulative frequency
0–4	10	
5–9	12	
10–14	7	
15–19	6	
20–24	5	
25–29	3	
30–34	2	

- a** Copy and complete the frequency distribution table.
b Calculate the median class for this distribution.
- 12** For the frequency distribution table in question **11**:
a make a list of the class centres for the distribution
b draw a cumulative frequency histogram and polygon
c use the cumulative frequency polygon to estimate the median of the distribution.
- 13** **WE12** For each of the following sets of scores find the mode.
a 2, 5, 3, 4, 5 **b** 8, 10, 7, 10, 9, 8, 8
c 11, 12, 11, 15, 14, 13 **d** 0.5, 0.4, 0.6, 0.3, 0.2, 0.4, 0.6, 0.9, 0.4
e 110, 113, 100, 112, 110, 113, 110
- 14** Find the mode for each of the following. (*Hint*: Some are bimodal and others have no mode.)
a 16, 17, 19, 15, 17, 19, 14, 16, 17 **b** 147, 151, 148, 150, 148, 152, 151
c 2, 3, 1, 9, 7, 6, 8 **d** 68, 72, 73, 72, 72, 71, 72, 68, 71, 68
e 2.6, 2.5, 2.9, 2.6, 2.4, 2.4, 2.3, 2.5, 2.6
- 15** **WE13** Use the tables below to state the mode of the distribution.

a

Score	Frequency
1	2
2	4
3	5
4	6
5	3

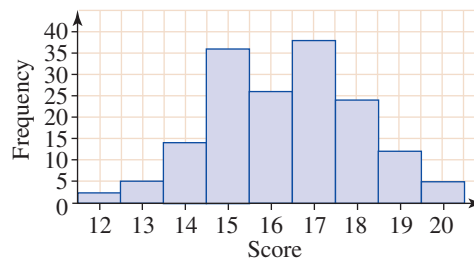
b

Score	Frequency
5	1
6	3
7	5
8	8
9	5
10	3

c

Score	Frequency
38	2
39	4
40	1
41	5
42	6
43	3
44	6
45	2

- 16** Use the frequency histogram below to state the mode of the distribution.



- 17** For each of the following grouped distributions, state the modal class.

a

Class	Frequency
1–4	6
5–8	12
9–12	30
13–16	23
17–20	46
21–24	27
25–28	9

b

Class	Frequency
1–7	3
8–14	8
15–21	9
22–28	25
29–35	12
36–42	11
43–49	2

- 18** The table at right shows the depth of snow during every day of the ski season.

- a** Redraw the table to include the class centres and cumulative frequency.
b Draw a cumulative frequency histogram and polygon.
c Use the graph to estimate the median depth of snow for the ski season.

Depth (cm)	Frequency
0–50	8
50–100	9
100–150	12
150–200	15
200–250	6
250–300	4
300–350	2
350–400	2

- 19** The weekly wage (in dollars) of 40 people is shown below.

376 592 299 501 375 366 204 359 382 274
 223 295 232 325 311 513 348 235 329 203
 556 419 226 494 205 307 417 204 528 487
 543 532 435 415 540 260 318 593 592 393

- a** Use the classes \$200–\$249, \$250–\$299, \$300–\$350 etc. to display the information in a frequency distribution table.
b From your table, calculate the median class.
c Draw a cumulative frequency histogram and polygon, and use it to estimate the median wage in the group.

Further development

- 20** Consider the stem and leaf plot below:

Stem	Leaf	
60	2 5 8	
61	1 3 3 6 7 8 9	
62	0 1 2 4 6 7 8 8 9	
63	2 2 4 5 7 8	key 60 3 = 603
64	3 6 7	
65	4 5 8	
66	3 5	
67	4	

- a** Find the median of the data set.
b Find the mode of the data set.
c Explain the advantage of a stem-and-leaf plot when trying to find the median and mode

- 21** Consider the following data set 23, 24, 20, 21, 25, 26, 28, 26.
- Find the median of the data set.
 - An extra score of 56 is added to the data set. Find the new median.
 - Describe the effect that the extra (outlier) score had on the median.
- 22** Will the addition of an outlier have any effect on the mode?
- 23** Consider the following set of 20 scores.
- 64, 34, 67, 22, 59, 72, 34, 93, 20, 82
30, 45, 27, 70, 44, 82, 71, 65, 45, 66
- Find the mode of the data set.
 - Daryl put the data into a table with a class size of 10 beginning with 20–29. Find the modal class.
 - Explain why the modal class will have more meaning than the mode.
- 24** Explain what will be meant by the term median class. Why do you think this term is seldom used?
- 25** When the median of a grouped distribution is found using an ogive the result will only be an estimate of the median. Explain.

10D Best summary statistics

Having now examined all three **summary statistics**, it is important to recognise when it is appropriate to use each one. In some circumstances, one summary statistic may be more appropriate than the others. For example, a shoe manufacturer notes that in a new style of sporting footwear:

mean size sold is 8.63
median size is 8.75
mode size is 9.

In this case, the mode is the most useful measure as the manufacturer needs to know which size sells the most. The mean and median are of less use to the manufacturer.

WORKED EXAMPLE 14

Below are the wages of ten employees in a small business.

\$420 \$430 \$490 \$475 \$465 \$450 \$1700 \$420 \$420 \$440

- Calculate the mean wage.
- Calculate the median wage.
- Calculate the mode wage.
- Does the mean, median or mode give the best measure of a typical wage in this business?

THINK

- Total all the wages.
 - Divide the total by 10.
- Write the wages in ascending order.
 - Average the 5th and 6th score to find the median.

WRITE

- Total = \$5710
Mean = $\$5710 \div 10$
= \$571
- $\$420 \ \$420 \ \$420 \ \$430 \ \$440 \ \$450 \ \$465$
 $\$475 \ \$490 \ \$1700$
 Median = $\frac{\$440 + \$450}{2}$
 = \$445

eBookplus

Tutorial
int-2328

Worked example 14

- c** \$420 is the score that occurs most often and so this is the mode.
- d** The mean is larger than what is typical because of one very large wage: the mode is the lowest wage and so this is not typical. Therefore, the median is the best measure.

c Mode = \$420

- d** The median is the best measure of the typical wage as the mode is the lowest score, which is not typical, and the mean is inflated by the \$1700 wage.

For each of these examples you will need to think carefully about the relevance of each summary statistic in terms of the particular example.

REMEMBER

- The three summary statistics are:
 mean — calculated by adding all scores, then dividing by the number of scores
 median — the middle score or average of the two middle scores
 mode — the score with the highest frequency.
- Be careful when using the mean. One or two extreme scores can greatly increase or decrease its value.
- When the mean is not a good measure of central tendency, the median is used.
- The mode is the best measure in some examples where discrete data means that the mean and median may have very little meaning.

EXERCISE

10D Best summary statistics

eBookplus

Digital doc

GC program —

Casio

doc-1592

UV stats

eBookplus

Digital doc

GC program — TI

doc-1593

UV stats

- 1 WE14** There are ten houses in a street. A real-estate agent values each house with the following results.

\$350 000 \$390 000 \$375 000 \$350 000 \$950 000
 \$350 000 \$365 000 \$380 000 \$360 000 \$380 000

- Calculate the mean house valuation.
- Calculate the median house valuation.
- Calculate the mode house valuation.
- Which of the above is the best measure of central tendency?

- 2** The table below shows the number of shoes of each size that were sold over a week at a shoe store.

Size	Frequency
4	5
5	7
6	19
7	24
8	16
9	8
10	7

- Calculate the mean shoe size sold.
- Calculate the median shoe size sold.
- Calculate the mode of the data set.
- Which measure of central tendency has the most meaning to the shoe store proprietor?



- 3 The table below shows the crowds at football matches over a season.

Crowd	Class centre	Frequency
10 000–20 000	15 000	95
20 000–30 000	25 000	64
30 000–40 000	35 000	22
40 000–50 000	45 000	15
50 000–60 000	55 000	3
60 000–70 000	65 000	0
70 000–80 000	75 000	1

- a Calculate the mean crowd over the season.
b Calculate the median class.
c Calculate the modal class.
d Draw a cumulative frequency histogram and polygon.
e Use the ogive to estimate the median.
f Which measure of central tendency would best describe the typical crowd at football matches over the season?
- 4 **MC** Mr and Mrs Yousef research the typical price of a large family car. At one car yard they find six family cars. Five of the cars are priced between \$30 000 and \$40 000, while the sixth is priced at \$80 000. What would be the best measure of the price of a typical family car?
- A Mean B Median C Mode D All are equally important.
- 5 Thirty men were asked to reveal the number of hours they spent doing housework each week. The results are given below.

1 5 2 12 2 6 2 8 14 18
0 1 1 8 20 25 3 0 1 2
7 10 12 1 5 1 18 0 2 2

- a Represent the data in a frequency distribution table. (Use classes 0–4, 5–9, 10–14 etc.)
b Find the mean number of hours that the men spend doing housework.
c Find the median class for hours spent by the men at housework.
d Find the modal class for hours spent by the men at housework.
- 6 The resting pulse rates of 20 female athletes were measured. The results are shown below.

50 62 48 52 71
61 30 45 42 48
43 47 51 52 34
61 44 54 38 40

- a Represent the data in a frequency distribution table using appropriate groupings.
b Find the mean of the data.
c Find the median class of the data.
d Find the modal class of the data.
e Draw an ogive of the data.
f Use the ogive to determine



the median pulse rate.

- 7** The following data give the age of 25 patients admitted to the emergency ward of a hospital.

18 16 6 75 24
23 82 74 25 21
43 19 84 72 31
74 24 20 63 79
80 20 23 17 19

- Represent the data in a frequency distribution table. (Use classes 0–14, 15–29, 30–44, etc.)
 - Use the table to find:
 - the mean age of patients admitted
 - the median class of age of patients admitted
 - the modal class for age of patients admitted.
 - Draw an ogive of the data.
 - Use the ogive to determine the median age.
 - Do any of your statistics (mean, median or mode) give a clear representation of the typical age of an emergency ward patient?
 - Give some reasons that could explain the pattern of the distribution of data in this question.
- 8** The batting scores for two cricket players over six innings are as follows:

Player A: 31, 34, 42, 28, 30, 41

Player B: 0, 0, 1, 0, 250, 0

- Find the mean score for each player.
 - Which player appears to be better if the mean result is used?
 - Find the median score for each player.
 - Which player appears to be better when the decision is based on the median result?
 - Which player do you think would be more useful to have in a cricket team and why? How can the mean result sometimes lead to a misleading conclusion?
- 9** The following frequency table gives the number of employees in different salary brackets for a small manufacturing plant.

Position	Salary (\$)	No. of employees
Machine operator	38 000	50
Machine mechanic	40 000	15
Floor steward	44 000	10
Manager	82 000	4
Chief Executive Officer	100 000	1

- Workers are arguing for a pay rise, but the management of the factory claims that workers are well paid because the mean salary of the factory is \$42 100. Are they being honest?
- Suppose that you were representing the factory workers and had to write a short submission in support of the pay rise. How could you explain the management's claim? Provide some other statistics to support your case.

Further development

- 10** The data below shows the age of 25 patients admitted to the emergency ward of a hospital.

18 6 16 75 24 23 82 75 25 21 43 19 84 76 30 78 24 20 63 79

- Find the mean age of the patients.

eBookplus

Digital doc
WorkSHEET 10.2
doc-1594

- b** Find the median age of the patients.
- c** What is the mode age?
- d** Do any of the measures of central tendency give a clear representation of the typical age of an emergency ward patient? Give a reason for your answer.

11 A small business pays the following salaries (in thousands of dollars) to its employees:

38, 38, 38, 38, 46, 46, 46, 55, 100 (the manager)

- a** What is the salary of most workers?
- b** What is the mean salary?
- c** What is the median salary?
- d** The union is negotiating a salary rise for the workers. What measure would be used by:
 - i** the union in negotiations
 - ii** the employer in negotiations.

Explain each answer.

12 The contents of 20 randomly selected boxes of matches were counted. The following data shows the number of matches in each box:

138, 139, 139, 141, 137, 140, 137, 141, 139, 142
140, 141, 141, 139, 141, 138, 139, 140, 141, 138



- a** Find the mean, median and mode of the distribution.
 - b** Which of the three measures best supports the manufacturers claim that there are 140 matches per box?
 - c** Is this claim by the manufacturer valid?
- 13** A class of mathematics students got a median mark of 54 for their end of semester test; however no-one actually scored this result.
- a** Explain how this is possible.
 - b** How many students must have scored below 54?
- 14**
- a** What is the effect of an outlier on the mean of a data set?
 - b** When this occurs what is usually the best measure of central tendency?
 - c** Give an example of when the mode is the most relevant measure of central tendency.
- 15**
- a** Explain why the range is an unreliable measure of spread.

- b** Does a single outlier have any effect on the interquartile range and standard deviation?

INVESTIGATE: Wage rise

The workers in an office are trying to obtain a wage rise. In the previous year, the ten people who work in the office received a 2% rise while the company CEO received a 42% rise.

- 1** What was the mean wage rise received in the office last year?
- 2** What was the median wage rise received in the office last year?
- 3** What was the modal wage rise received in the office last year?
- 4** The company is trying to avoid paying the rise. What statistic do you think they would quote about last year's wage rises? Why?
- 5** What statistic do you think the trade union would quote about wage rises? Why?
- 6** Which statistic do you think is the most 'honest' reflection of last year's wage rises? Explain your answer.

Quoting different averages can give different impressions about what is normal. Try the following task.

- 1** Visit a local real estate agent and study the properties for sale in the window.
 - 2** Calculate the mean, median and mode price for houses in the area.
 - 3** If you were a real estate agent and a person wanting to sell their home asked what the typical property sold for in the area, which figure would you quote?
 - 4** Which figure would you quote to a person who wanted to buy a house in the area?
-

INVESTIGATE: Best summary statistics and comparison of samples

Examine each of the following statistics.

- The typical mark in Maths among Year 11 students.
 - The number of attempts taken by Year 11 and 12 students to get their driver's licence.
 - The typical number of days taken off school by Year 11 students so far this year.
- 1** For each of the above, gather your data by selecting a random sample.
 - 2** Calculate the mean, median and mode for each topic.
 - 3** Compare your results with other students who will have selected their samples from the same population.
 - 4** In each case, state the best summary statistic and explain your answer.
-

SUMMARY

Calculating the mean

- For a small number of scores, the mean is calculated using the formula:

$$\bar{x} = \frac{\sum x}{n}.$$

- When the data are presented in a frequency table, the mean can be calculated using the formula:

$$\bar{x} = \frac{\sum f \times x}{\sum f}.$$

- The mean can also be calculated using the statistical function on your calculator.

Standard deviation

- The standard deviation is a measure of the spread of a data set.
- The smaller the standard deviation, the smaller the spread of the data set.
- The standard deviation is found using the statistical function on your calculator.
- When the analysis is conducted on the entire population, the population standard deviation (σ_n) is used.
- When the analysis is conducted on a sample of the population, the sample standard deviation (σ_{n-1}) is used.

Median and mode

- The median is the middle score of a data set, or the average of the two middle scores.
- The mode is the score with the highest frequency.

Best summary statistics

- The summary statistics are the mean, median and mode.
- Each summary statistic must be examined in the context of the statistical analysis to determine which is the most relevant.

CHAPTER REVIEW

MULTIPLE CHOICE

- 1 **MC** For the following data set, which of the statements is correct?
3, 4, 8, 7, 3, 6, 5, 3, 4, 7
 - A The mean is 5.
 - B The median is 5.
 - C The mode is 5.
 - D all of the above
- 2 **MC** For which of the following data sets is the median greater than the mean?
 - A 2, 6, 14, 14, 15, 16, 18
 - B 12, 13, 14, 14, 14, 18, 22
 - C 12, 15, 15, 15, 15, 18
 - D 1, 4, 9, 16, 25, 36, 49
- 3 **MC** For the data set below, which statement is correct?
25, 45, 64, 48, 66, 85, 45, 27
 - A The mean is 50.625.
 - B The sample standard deviation is 20.29.
 - C The population standard deviation is 18.98.
 - D all of the above
- 4 **MC** Tracey compiles a sample of new car prices. She selects 100 new car buyers and asks what price they paid for their car. To measure the spread of the distribution Tracey should use:
 - A the population standard deviation
 - B the sample standard deviation
 - C both standard deviations
 - D the mean
- 5 **MC** For the statistical analysis in question 4 which summary statistic would be the most appropriate?
 - A mean
 - B median
 - C mode
 - D standard deviation

SHORT ANSWER

- 1 Calculate the mean of each of the following sets of scores.
 - a 4, 9, 5, 3, 5, 6, 2, 7, 1, 10
 - b 65, 67, 87, 45, 90, 92, 50, 23
 - c 7.2, 7.9, 7.0, 8.1, 7.5, 7.5, 8.7
 - d 5, 114, 23, 12, 25

- 2 Copy and complete the tables below and then use them to calculate the mean.

a

Score (x)	Frequency (f)	$f \times x$
5	11	
6	15	
7	24	
8	21	
9	9	
$\Sigma f =$		$\Sigma f \times x =$

b

Score (x)	Frequency (f)	$f \times x$
9.2	36	
9.3	48	
9.4	74	
9.5	65	
9.6	51	
9.7	32	
9.8	14	
9.9	2	
$\Sigma f =$		$\Sigma f \times x =$

- 3 Complete the frequency distribution table below and use it to estimate the mean of the distribution.

Class	Class centre (x)	Frequency (f)	$f \times x$
21–24		3	
25–28		9	
29–32		17	
33–36		31	
37–40		29	
41–44		25	
45–48		19	
49–52		10	
		$\Sigma f =$	$\Sigma f \times x =$

- 4 Use the statistics function on your calculator to find the mean of each of the following sets of scores.
 - a 2, 18, 26, 121, 96, 32, 14, 2, 0, 0
 - b 2, 2, 12, 12, 12, 32, 32, 47, 58
 - c 0.2, 0.3, 0.6, 0.4, 0.3, 0.7, 0.8, 0.6, 0.5, 0.4, 0.1

- 5 Use the statistics function on your calculator to find the mean of the following distributions. Where necessary, give your answers correct to 1 decimal place.

a

Score	Frequency
10	23
20	47
30	68
40	56
50	17

b

Score	Frequency
24	45
25	89
26	124
27	102
28	78
29	46

c

Class	Class centre	Frequency
10–12	11	18
13–15	14	32
16–18	17	34
19–21	20	40
22–24	23	28
25–27	26	14
28–30	29	6

- 6 The marks of 30 students in a Geography test are shown below.

66 47 43 80 42 92 92 90 92 77
 67 87 75 72 42 60 86 53 95 78
 46 87 49 70 82 92 93 71 62 67

- a** Calculate the mean.
b Should the population or sample standard deviation be used in this case?
c Write the value of the appropriate standard deviation.
- 7 To find the number of attempts most people take to get their driver's licence, a sample of twenty Year 12 students is chosen. The results are shown below.

1 2 3 3 1 2 1 2 4 1
 1 1 2 2 2 3 1 2 2 3

- a** Calculate the mean.
b Should the population or sample standard deviation be used in this case?
c Write the value of the appropriate standard deviation.

- 8 Use the statistics function on your calculator to find the mean and population standard deviation of each of the following distributions. Give each answer correct to 3 decimal places.

- a** 0.7, 1.2, 0.5, 0.9, 1.3, 1.5, 0.1, 1.0, 0.4, 0.5
b 23, 254, 12, 89, 74, 15, 26, 45

c

Score	Frequency
26	12
27	25
28	29
29	28
30	14

d

Class	Class centre	Frequency
10–14	12	8
15–19	17	12
20–24	22	32
25–29	27	45
30–34	32	40
35–39	37	19
40–44	42	6

- 9 For each of the following sets of scores, find the median.

- a** 25, 26, 26, 27, 27, 28, 30, 32, 35
b 4, 5, 8, 5, 8, 6, 7, 10, 4, 8, 4
c 3.2, 3.1, 3.0, 3.5, 3.2, 3.2, 3.2, 3.6
d 2, 3, 7, 4, 4, 8, 5, 7, 7, 6
e 121, 135, 111, 154, 147, 165, 101, 108

- 10 Copy and complete each of the following frequency tables and then use them to find the median.

a

Score	Frequency	Cumulative frequency
0	2	
1	6	
2	11	
3	7	
4	6	
5	3	

b

Score	Frequency	Cumulative frequency
54	2	
55	5	
56	14	
57	11	
58	6	
59	1	
60	1	

c

Score	Frequency	Cumulative frequency
66	8	
67	10	
68	12	
69	14	
70	7	
71	5	
72	4	

- 11 a** Copy and complete the frequency distribution table below.

Class	Class centre	Frequency	Cumulative frequency
30–39		18	
40–49		34	
50–59		39	
60–69		45	
70–79		29	
80–89		10	
90–99		5	

- b** What is the median class of this distribution?
c Display these data in a cumulative frequency histogram and polygon.
d Use your graph to estimate the median of the distribution.
- 12** For each set of scores below, state the mode.
a 2, 3, 6, 8, 4, 2, 4, 2, 6, 5, 2
b 23, 24, 19, 23, 27, 25, 31, 24, 23, 27, 27
c 1.2, 5.6, 4.7, 6.8, 4.5, 2.1

- 13** For each of the frequency tables below, state the mode.

a

Score	Frequency
1	23
2	35
3	21
4	19
5	8

b

Score	Frequency
14	9
15	15
16	8
17	12
18	15
19	7
20	1

- 14** Use the frequency table below to state the modal class.

Class	Class centre	Frequency
30–33	31.5	12
34–37	35.5	26
38–41	39.5	34
42–45	43.5	45
46–49	47.5	52
50–53	51.5	23

- 15** Below are the number of goals scored by a netball team in ten matches in a tournament.

25 26 19 24 28 67 21 22 28 18

- a** Calculate the mean.
b Calculate the median.
c Calculate the mode.
d Which of the above is the best summary statistic? Explain your answer.
- 16** Give an example of a statistical analysis where the best summary statistic is:
a the mean
b the median
c the mode.

EXTENDED RESPONSE

- 1 The table below shows the gross annual income for a sample of 100 company executives.

Income	Class centre	Frequency	Cumulative frequency
\$50 000–\$75 000		12	
\$75 000–\$100 000		18	
\$100 000–\$125 000		26	
\$125 000–\$150 000		24	
\$150 000–\$175 000		12	
\$175 000–\$200 000		8	

- Copy and complete the frequency table.
 - Calculate the mean.
 - Calculate the standard deviation.
 - Calculate the median class
 - Calculate the modal class.
 - Which summary statistic best describes the typical income for a company executive?
- 2 In order to compare two textbooks, a teacher recommends one book to one class and another book to another class. At the end of the year the classes are each tested; the results are detailed below.

Text A

44 52 95 76 13 94 83 72 55 81 22 25 64
72 35 48 56 59 84 98 84 21 35 69 28

Text B

65 72 48 63 68 59 68 62 75 79 81 72 64 53
58 59 64 66 68 42 37 39 55 58 52 82 79 55

- Calculate the mean and standard deviation for each class group.
- Which standard deviation did you use in part a? Explain why.
- Which class performed better?
- Which class had the more consistent results?
- Could a conclusion be drawn about the better textbook from the above information? Explain your answer.

eBook^{plus}

Digital doc
Test Yourself
doc-1595
Chapter 10

Are you ready?**Digital docs** (page 324)

- SkillsSHEET 10.1 (doc-1581): Finding the mean of a list of scores
- SkillsSHEET 10.2 (doc-1582): Presenting data as a dot plot
- SkillsSHEET 10.3 (doc-1583): Presenting data in a frequency table
- SkillsSHEET 10.5 (doc-1585): Presenting data as a stem-and-leaf plot

10A Calculating the mean**Tutorial**

- **WE3** int-2327: Learn to construct a frequency distribution for a set of data. (page 327)

Digital docs

- SkillsSHEET 10.1 (doc-1581): Finding the mean of a list of scores (page 329)
- SkillsSHEET 10.2 (doc-1582): Presenting data as a dot plot (page 329)
- SkillsSHEET 10.3 (doc-1583): Presenting data in a frequency table (page 329)
- SkillsSHEET 10.4 (doc-1584): Organising data into class intervals (page 330)
- SkillsSHEET 10.5 (doc-1585): Presenting data as a stem-and-leaf plot (page 330)
- Spreadsheet (doc-1586): Mean (page 330)
- Spreadsheet (doc-1587): Mean (DIY) (page 330)

10B Standard deviation**Interactivity**

- int-2402: Standard deviation (page 334)

Digital docs

- WorkSHEET 10.1 (doc-2484): Apply statistical measures to questions (page 340)

10C Median and mode**Digital docs**

- Spreadsheet (doc-1588): Median (page 345)
- Spreadsheet (doc-1589): Median (DIY) (page 345)
- Spreadsheet (doc-1590): Mode (page 345)
- Spreadsheet (doc-1591): Mode (DIY) (page 345)

10D Best summary statistics**Tutorial**

- **WE14** int-2328: Learn to distinguish between the best summary statistics. (page 349)

Digital docs

- GC program — Casio (doc-1592): UV stats (page 350)
- GC program — TI (doc-1593): UV stats (page 350)
- WorkSHEET 10.2 (doc-1594): Apply your knowledge of statistics to questions. (page 352)

Chapter review

- Test yourself Chapter 10 (doc-1595): Take the end-of-chapter test to test your progress. (page 359)

To access eBookPLUS activities, log on to

www.jacplus.com.au